Polygenic Risk Scores for Cardio-renal-metabolic Diseases in the Penn Medicine Biobank

Kember, R.L.[1,*], Verma, A.[1,*], Verma, S.[1,*], Lucas, A.[1], Judy, R.[2], Chen, J.[3], Damrauer, S.[2], Rader, D.J.[1], Ritchie, M.D.[1,**]

[1]Department of Genetics, Perlman School of Medicine, University of Pennsylvania, Philadelphia PA, USA
[2]Department of Surgery, Perlman School of Medicine, University of Pennsylvania, Philadelphia PA, USA
[3]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, Philadelphia, PA, USA

*These authors contributed equally to this work

**Correspondence email: marylyn@upenn.edu

Abstract

Cardio-renal-metabolic (CaReMe) conditions are common and the leading cause of mortality around the world. Genome-wide association studies have shown that these diseases are polygenic and share many genetic risk factors. Identifying individuals at high genetic risk will allow us to target prevention and treatment strategies. Polygenic risk scores (PRS) are aggregate weighted counts that can demonstrate an individual's genetic liability for disease. However, current PRS are often based on European ancestry individuals, limiting the implementation of precision medicine efforts in diverse populations. In this study, we develop PRS for six diseases and traits related to cardio-renal-metabolic disease in the Penn Medicine Biobank. We investigate their performance in both European and African ancestry individuals, and identify genetic and phenotypic overlap within these conditions. We find that genetic risk is associated with the primary phenotype in both ancestries, but this does not translate into a model of predictive value in African ancestry individuals. We conclude that future research should prioritize genetic studies in diverse ancestries in order to address this disparity.

Introduction

In this era of precision medicine, there are significant efforts to identify the genetic, environmental, family history, and clinical factors that influence the risk of disease as well as the influence of these factors on disease prognosis and treatment. Knowing in advance the factors that can lead to increased risk of disease can provide a major health benefit to individuals, as treatment and support strategies can be targeted towards individuals at higher risk. Identification of a large number of loci with small genetic effects in genome-wide association studies (GWAS) have highlighted the polygenic behavior of most common, complex diseases[1,2]. An emerging technology in the field of disease risk prediction is the polygenic risk score (PRS). PRS is the cumulative, mathematical aggregation of risk derived from the contributions of many DNA variants across the genome[3].

Recent studies show high prevalence of cardio-renal-metabolic conditions among adults in the USA[4] and together they are the leading cause of mortality around the world[5,6]. GWAS have

identified more than 100 loci associated with common diseases such as coronary artery disease (CAD), body mass index (BMI), hypertension, renal failure and type 2 diabetes (T2D). This group of cardio, renal, and metabolic conditions are collectively referred to as CaReMe conditions. Among the individuals that are diagnosed with one disease (for example T2D), the prevalence of comorbidities such as hypertension, CAD, heart failure (HF), and chronic kidney disease (CKD) also increases. To evaluate disease risk in an individual, it is essential to also consider comorbid or secondary conditions related to the primary disease. There are several GWA-studies that have identified shared genetic associations between CaReMe conditions, demonstrating similarity in the underlying genetic architecture[7,8]. Pathophysiology of these conditions also show the cross-talk between organ systems and its effect on disease progression such as hemodynamic interaction between heart and kidney in heart failure[9]. With PRS, we can derive individuals' disease risk for each CaReMe condition using GWAS summary statistics. More importantly, PRS is derived from the effect of millions of genetic variants on a disease; so it accounts for an individual's genetic background. Therefore, PRS can evaluate the genetic overlap among coexisting or comorbid conditions. Phenome-wide Association Studies (PheWAS) can be used to identify links between disease risk and other conditions[10–12]. Using these strategies, we investigated whether cross-phenotype associations can provide insights into the contribution of risk for one disease risk on other conditions. Lastly, we also evaluated the effect of age, sex, and ancestry on CaReMe PRS predictions.

There are several strategies to derive PRS for a disease of interest. Traditionally, genetic risk scores (GRS) were derived using the genome-wide significant SNPs from a genome-wide association study; however, recent studies show that using association results with much lower p-value significance ($p<0.05$) segregate individuals risk with better accuracy[1]. The development and clinical utility of PRS is under active investigation, especially in racial and ethnic minority populations[13–15]. Most large-scale GWAS have been conducted in individuals from European descent populations and most PRS are derived from these studies. Subsequently, the majority of PRS investigations published to date have been conducted in populations of European ancestry[16]. There can be several differences such as linkage disequilibrium (LD) structure and allele frequency of the variants, which can lead to inaccurate PRS for non-European populations[16]. This is not unique to PRS studies, but the majority of human genetic research suffers from this same phenomenon[17]. In this study, we investigated the implementation of PRS for cardio-renal-metabolic conditions in European (EUR) and African (AFR) ancestry

individuals in the Penn Medicine Biobank (PMBB). PMBB is a cohort of 52,853 individuals established for genomic and precision medicine research. Approximately 20,000 of the individuals in the study have genetic data from a genotyping array which has been imputed to the 1000Genomes phase III using the Michigan Imputation Server[18]. Approximately 25% of the PMBB study population is African ancestry. We calculated PRS in the PMBB genetic data to evaluate 1) risk prediction accuracy among EUR and AFR based on GWAS summary statistics generated in EUR data; and 2) the utility of PRS in determining genetic overlap among CaReMe conditions.

Methods

Penn Medicine Biobank

The Penn Medicine BioBank (PMBB) recruits participants through the University of Pennsylvania Health System by enrolling at the time of appointment. Patients participate by donating either blood or a tissue sample and allowing researchers access to their electronic health record (EHR) information. This academic biobank provides researchers with centralized access to a large number of blood and tissue samples with attached health information. The facility banks both blood specimens (i.e., whole blood, plasma, serum, buffy coat, and DNA isolated from leukocytes) and tissues (i.e., formalin-fixed paraffin-embedded, fresh and flash frozen). PMBB currently consists of 52,853 consented samples. Approximately one third (N=19,515) of these participants have been genotyped to date. PMBB is a diverse cohort, with 70% European ancestry, 25% African ancestry, and 5% Asian or Latino ancestry. See Table 1 for characteristics of all participants.

Table 1. Participant Characteristics

	PMBB consented patients	**PMBB genotyped patients**
Total Patients	52,853	19,515
Female, (%)	25,926 (49%)	7856 (41%)
Age (average)	18 – 99 (60)	20 – 99 (66)

Body mass index	29.36 (13 - 83)	30.02 (13 - 83)
Race		
American Indian or Alaska Native	34	1
Asian	1,201	158
Black or African American	11,173	6159
Native Hawaiian or other Pacific Islander	30	7
Other	1,577	459
Unknown	1,775	533
White	36,707	10,563
Ethnicity		
Hispanic or Latino	1,381	350
Not Hispanic or Latino	50,994	17,517
Unknown	122	13

Genotyping and Quality Control and Imputation

DNA extracted from the blood plasma of 19,515 samples were genotyped in three batches: 10,867 samples on the Illumina QuadOmni chip at the Regeneron Genetics Center; 5,676 samples on the Illumina GSA V1 chip and 2,972 samples on the Illumina GSA V2 chip by the Center for Applied Genomics at the Children's Hospital of Philadelphia. Due to the low overlap among genetic variants on the different genotyping arrays, we used an imputation strategy to combine these datasets[18,19]. Prior to imputation, we applied a quality control pipeline[19] to each dataset, removing individuals with sex errors or had a sample call rate <90%; and removing variants which were palindromic or had a call rate <95%. Table 2 summarizes each dataset before and after QC.

Table 2: Summary of genotype data before and after QC.

Dataset	Pre-QC		Post-QC	
	#Samples	#SNPs	#Samples	#SNPs
Illumina Infinium OMNI	10,867	713,599	10,506	651,366
Illumina GSA V1	5,676	700,078	5,660	666,032
Illumina GSA V2	2,972	759,993	2,965	700,984

Genotypes for each of the three PMBB datasets were phased (Eagle v2.3) and imputed to the 1000Genomes reference panel (1000G Phase3 v5) using the Michigan Imputation Server[18]. Accuracy of the imputed variants was assessed via comparison of the expected vs actual allele frequency of variants (R^2=0.3). Following imputation, the datasets were merged, with each position matched based on alleles. In the merged dataset, the average R^2 of variants = 0.75. Genetic ancestry was calculated from common, high-quality SNPs (MAF > 0.05, missingness < 0.1) using SMARTPCA[20] module of the Eigensoft package. We split the merged file into individuals with European ancestry (N=11,524) and individuals with African ancestry (N=5,994). All subsequent QC and analysis steps were performed independently within each population.

We retained high quality, common SNPs with imputation marker R2 ≥ 0.7 and minor allele frequencies ≥ 0.01. We identified and removed related individuals using a kinship coefficient of 0.25. Using a graph-based algorithm, we selected and removed the sample that is closely related to the most samples within the set of related samples. Following QC, we retained 10,351 European ancestry individuals and 5,553 African ancestry individuals. Ancestry specific principal components were generated within each ancestral group following ancestry assignment, and these were used as covariates for subsequent analyses. Genetic ancestry of individuals was determined by performing quantitative discriminate analyses on PCs.

Polygenic Risk Scores

To derive PRS, we used the summary statistics from the largest and/or most recent GWAS studies for each trait (See Table 3). To reduce our total SNP set to a size amenable for PRS analysis, we first extracted SNPs present in the HapMap reference panel (N SNPs =1,437,731 in HapMap panel; retained 1,320,405 SNPs in EUR and AFR datasets). LDPred (v1.0) was used to generate the posterior mean effect of each SNP based on the LD information from either the PMBB European dataset or the PMBB African dataset[21]. PRS on each population were calculated using PLINK v1.9[22]. We tested several values for LDPred's tuning parameter "fraction of causal variants" (ρ=0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1) for deriving SNP weights.

Table 3: Source of GWAS used for PRS creation

Disease	GWAS source	Sample size in GWAS	Reference
Type 2 Diabetes	DIAGRAM	26,488 cases	PMID: 24509480
Body Mass Index	GIANT	339,224	PMID: 25673413
Hypertension	UK Biobank	144,793	PMID: 30940143
Myocardial Infarction	CARDIoGRAMplus C4D consortium	43,154 cases	PMID: 26343387
Coronary Artery Disease	CARDIoGRAMplus C4D consortium	60,801 cases	PMID: 26343387
Chronic Kidney Disease	CKD Genetics consortium	41,395 cases	PMID: 31152163

Phenotypes

We derived phenotypes using ICD-9 and ICD-10 data for 52,853 individuals from the electronic health record (EHR), consisting of 11.8 million records. We filtered on encounter type

to identify records representing encounters with a physician (see Supplemental Table 1 for encounters selected). ICD-9 codes were aggregated to phecodes using the phecode ICD9 map 1.2[10,23]; ICD-10 codes were aggregated to phecodes using the phecode ICD-10cm map 1.2 (beta)[24]. Individuals are considered cases for the phenotype if they have at least 2 instances of the phecode on unique dates, controls if they have no instance of the phecode, and 'other/missing' if they have one instance or a related phecode. By the following criteria, there were a total of 1,812 phecodes included in the analysis.

Statistical Analysis

PRS were standardized with mean = 0 and SD = 1. Logistic regression models accounting for age, sex, and the first 10 within-ancestry principal components (PCs) were used to test for association of PRS with the primary phenotype. Area under the receiver operator curve (AUC) was determined using the R package pROC, using the same logistic regression model as above. AUC was also calculated for covariates alone.

A Phenome-wide Association Study (PheWAS) was performed for the optimal PRS identified in the above analysis for each primary condition. Logistic regression models with each PRS as the independent variable, phecodes as the dependent variables, and age, sex, and the first 10 PCs as covariates were used to identify secondary phenotypic associations. A phenome-wide bonferroni significance threshold of 2.7×10^{-5} was applied to account for multiple testing.

Results

Demographics of PMBB dataset

Using phecodes, we identified 7,476 EUR ancestry individuals (73.4%) and 4,177 AFR ancestry individuals (76.4%) with either type 2 diabetes, obesity, hypertension, myocardial infarction, coronary atherosclerosis, or renal failure (Table 4). In EUR ancestry individuals, 24.7% had been diagnosed with one instance of disease, 35.2% had 2-3 diseases, 12.5% had 4-5 diseases, and 1.0% had all six diseases. In AFR ancestry individuals, 20.8% had been diagnosed

with one instance of disease, 39.1% had 2-3 diseases, 14.6% had 4-5 diseases, and 1.9% had all six diseases.

Table 4: Counts for phecodes in the Penn Medicine Biobank

Phecode	**Description**	**# EUR (total=10,182)**	**# AFR (total=5,465)**
250.2	Type 2 diabetes	2,119	1,804
278.1	Obesity	1,720	2,107
401	Hypertension	5,980	3,455
411.2	Myocardial Infarction	1,193	469
411.4	Coronary atherosclerosis	3,996	975
585	Renal failure	2,191	1,635

Determining the PRS with the best discriminative capacity

We generated a PRS for each phenotype of interest: type 2 diabetes, body mass index, hypertension, myocardial infarction, coronary artery disease, and chronic kidney disease (see methods, Table 3). Candidate PRS were generated for 7 parameters, and their association with the primary phenotype tested. All PRS had at least one parameter that was significantly associated with their primary phenotype (Figure 1, Supplemental Table 2). We selected the best performing PRS based on the maximum area under the receiver operator curve (AUC; Supplemental Table 3). Type 2 diabetes PRS was significantly associated with type 2 diabetes (best parameter for EUR ρ=0.01, OR=1.52, p=6.62x10^{-43}, best parameter for AFR ρ=0.01, OR=1.3, p=2.19x10^{-13}). BMI PRS was significantly associated with obesity (best parameter for EUR ρ=0.3, OR=1.7, p=8.97x10^{-65}, best parameter for AFR ρ=0.1, OR=1.2, p=5.55x10^{-10}). Hypertension PRS was significantly associated with hypertension (best parameter for EUR ρ=1,

OR=1.4, p=4.50x10^{-40}, best parameter for AFR ρ=0.3, OR=1.27, p=4.31x10^{-10}). Myocardial infarction PRS was significantly associated with myocardial infarction (best parameter for EUR ρ=0.01, OR=1.8, p=3.74x10^{-51}, best parameter for AFR ρ=0.1, OR=1.3, p=5.01x10^{-7}). Coronary artery disease PRS was significantly associated with coronary atherosclerosis (best parameter for EUR ρ=0.01, OR=1.66, p=1.54x10^{-77}, best parameter for AFR ρ=0.03, OR=1.27, p=2.53x10^{-8}). Chronic kidney disease PRS was significantly associated with renal failure (best parameter for EUR ρ=0.01, OR=1.2, p=2.60x10^{-6}, best parameter for AFR ρ=0.001, OR=1.1, p=0.024).

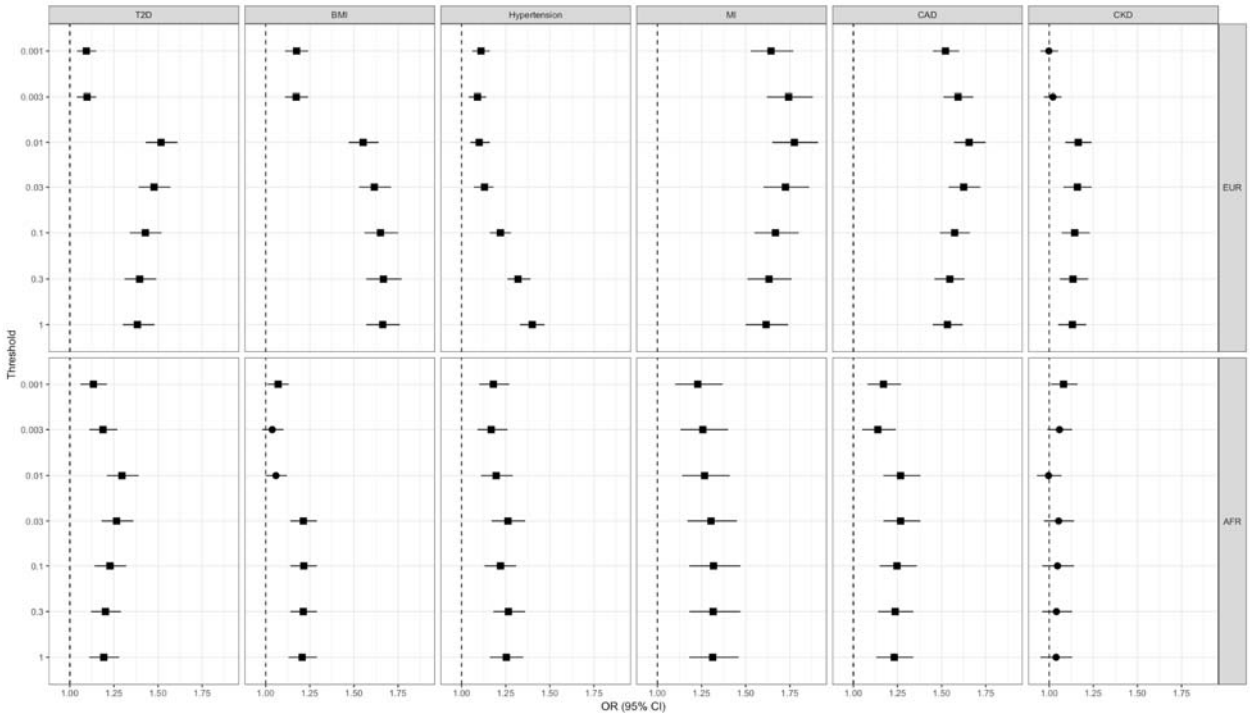

Figure 1. Association of PRS with primary phenotype for EUR and AFR ancestry. Square denotes p<0.05.

Performance of PRS in EUR vs. AFR ancestry

We selected the parameter that produced the strongest associated candidate PRS for each PRS-phenotype grouping within each ancestry for further analysis. In all cases, PRS performance was best in European ancestry individuals. The distribution of PRS in cases and controls in both

populations is illustrated in Figure 2. In European ancestry, the mean distribution of PRS in cases is consistently higher than controls. In African ancestry individuals this difference is much smaller, with substantial overlap between the PRS distribution in cases and controls. This is reflected in the comparison between the AUC for the full model and the AUC for covariates alone (Supplemental Table 3). Although the AUC in the full model is high in both ancestries (0.57-0.84), showing ability to distinguish between cases and controls, in AFR the full model offers little improvement over the model based on covariates alone (average improvement in AUC for best PRS=0.007). In contrast, in EUR the covariate model is improved when the PRS is added (average improvement in AUC for best PRS=0.032). Further, to evaluate the significance of variables in the full model (PRS and covariates), we performed a step-wise regression. In AFR, we identified that PRS was not selected in the best model in CAD but was selected in other phenotypes. CAD step-wide regression model further support our argument that PRS derived from EUR GWAS studies might not add any risk prediction in AFR individuals (Supplementary Table 4).

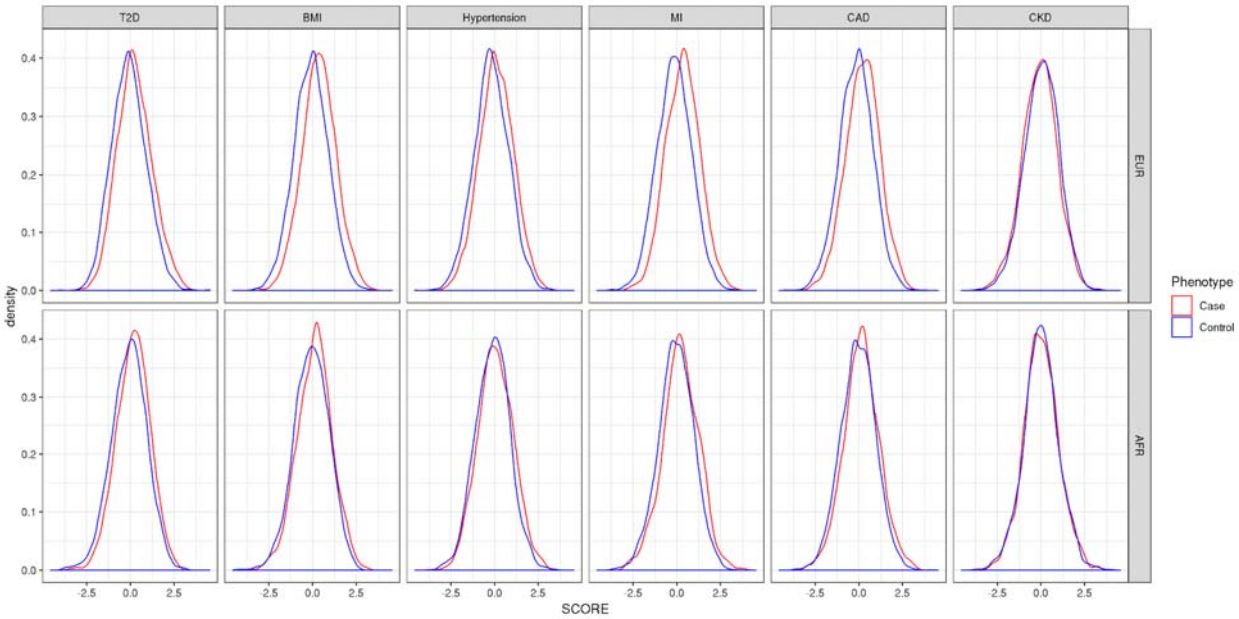

Figure 2. Density plot of polygenic risk scores in cases (red) and controls (blue) in both European and African ancestry individuals.

Correlation of PRS

We next tested for correlations between the polygenic risk scores for the six selected diseases in each ancestry (Figure 3, Supplemental Table 5). In European ancestry individuals, we identified significant positive correlations between CKD/T2D, CKD/BMI, CAD/BMI, CAD/Hypertension, CAD/MI, MI/BMI, MI/Hypertension, Hypertension/T2D, Hypertension/BMI, and BMI/T2D. CKD PRS was also negatively correlated with PRS for MI and CAD. In African ancestry individuals, we identified significant positive correlations between CKD/BMI, CAD/T2D, CAD/BMI, CAD/Hypertension, CAD/MI, MI/T2D, MI/BMI, MI/Hypertension, Hypertension/T2D, Hypertension/BMI, and BMI/T2D. Overall, the correlations identified in PRS in African ancestry individuals were weaker than those identified in PRS in European ancestry individuals.

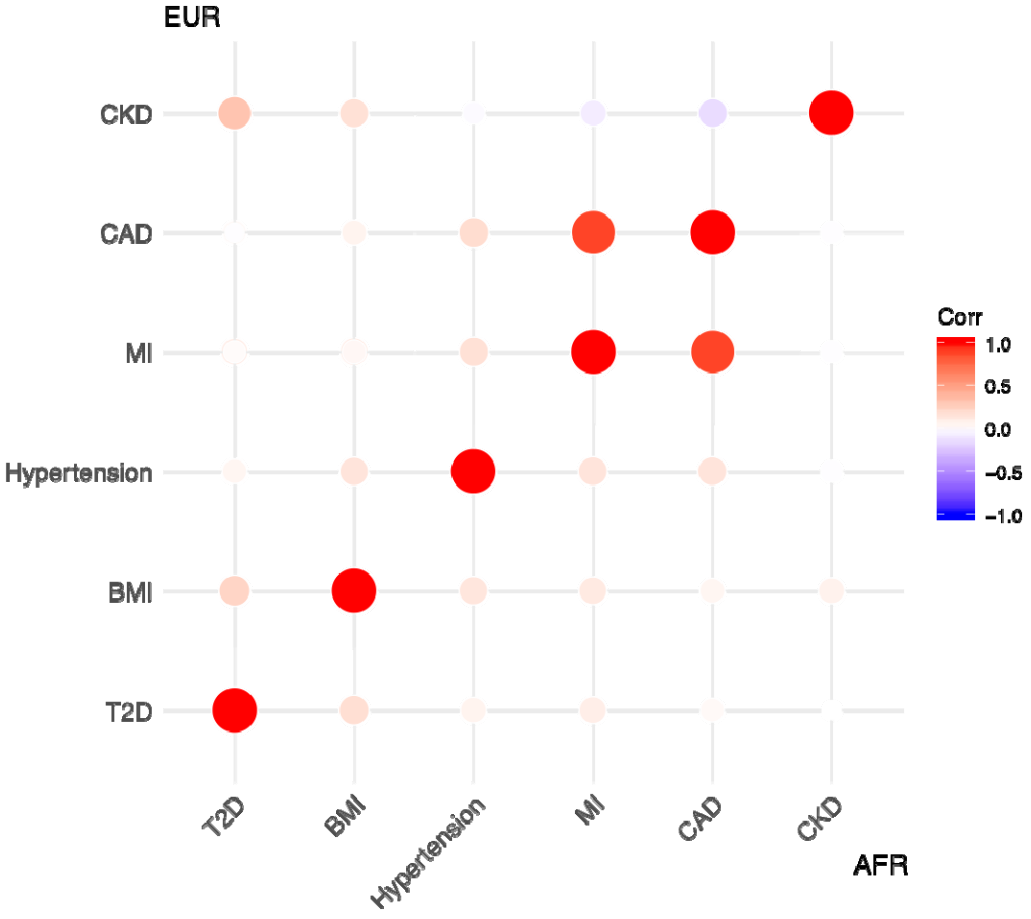

Figure 3. Correlation between polygenic risk scores for six diseases in European and African ancestry individuals. The size of the circle denotes the size of the correlation. The color denotes

the direction of correlation (red=positive, blue=negative). White circles denote a non-significant association. Correlation of European PRS are shown in the upper left, correlation of African PRS are shown in the lower right.

Association of PRS with disease burden in cases

Using linear regression in cases only, we tested whether increased PRS is associated with increased burden of cardio-renal-metabolic disease. All PRS, except for CKD PRS, were significantly associated with increased disease burden in both European and African ancestry individuals (Table 5). In Europeans, BMI PRS was the strongest association with occurrence of multiple diseases (beta=0.14, $p=6.6x10^{-19}$), in contrast to in African ancestry individuals, where the association remained significant but was reduced in effect size (beta=0.05, p=0.02). In African ancestry individuals, MI PRS was the strongest association with increased disease burden (beta=0.09, $p=8.3x10^{-7}$).

Table 5. Association of PRS with the burden of disease in European and African ancestry individuals.

PRS	**EUR**			**AFR**		
	Beta	**SE**	**P**	**Beta**	**SE**	**P**
T2D	0.117	0.017	$5.47x10^{-12}$	0.090	0.020	$7.18x10^{-6}$
BMI	0.135	0.015	$6.56x10^{-19}$	0.046	0.020	0.020
Hypertension	0.112	0.015	$2.40x10^{-14}$	0.037	0.019	0.048
MI	0.115	0.015	$3.85x10^{-15}$	0.093	0.019	$8.28x10^{-7}$
CAD	0.131	0.015	$6.31x10^{-18}$	0.092	0.019	$1.10x10^{-6}$
CKD	0.027	0.019	0.146	0.020	0.019	0.300

PheWAS of polygenic risk scores reveals secondary associations

We performed a PheWAS of each PRS in both European and African ancestry individuals to identify secondary phenotypes associated with genetic risk (Figure 4, Supplemental Tables 6-11). In all PRS, associations with secondary phenotypes in AFR individuals were reduced. All PRS except CKD PRS were associated with secondary phenotypes. T2D PRS was associated with hypertension (OR=1.13, $p=9.4x10^{-6}$) in EUR, and with Type 1 diabetes (OR=1.45, $p=4.8x10^{-6}$) in AFR. BMI PRS was associated with multiple circulatory system phenotypes in EUR, including hypertension (OR=1.17, $p=1.4x10^{-9}$), heart failure (OR=1.13, $p=1.1x10^{-6}$), and coronary atherosclerosis (OR=1.12, $p=5.3x10^{-6}$). BMI PRS was also associated with type 2 diabetes (OR=1.23, $p=1.2x10^{-14}$), renal failure (OR=1.13, $p=3.1x10^{-6}$), osteoporosis (OR=0.81, $p=5.3x10^{-5}$), and sleep apnea (OR=1.26, $p=4.3x10^{-15}$) in EUR. In AFR, the only additional phenotypes associated with BMI PRS were sleep apnea (OR=1.20, $p=1.1x10^{-6}$) and use of an insulin pump (OR=1.28, $p=8.6x10^{-6}$). Hypertension PRS was associated with circulatory system phenotypes in EUR, as well as type 2 diabetes (OR=1.29, $p=6.8x10^{-23}$), disorders of lipoid metabolism (OR=1.16, $p=2.6x10^{-10}$), renal failure (OR=1.13, $p=3.2x10^{-6}$), and sleep apnea (OR=1.15, $p=3.7x10^{-7}$). In AFR, hypertension PRS was not significantly associated with additional phenotypes. MI PRS was associated with hyperlipidemia (OR=1.22, $p=1.8x10^{-16}$), hypertensive chronic kidney disease (OR=1.26, $p=5.2x10^{-9}$) and type 2 diabetes (OR=1.12, $p=9.4x10^{-6}$) in EUR, but no additional phenotypes in AFR. CAD PRS is associated with hyperlipidemia (OR=1.29, $p=3.1x10^{-24}$), hypertensive chronic kidney disease (OR=1.28, $p=1.4x10^{-9}$), hypertension (OR=1.16, $p=6.3x10^{-9}$), type 2 diabetes (OR=1.14, $p=9.6x10^{-7}$), and disorders of eye (OR=0.77, $p=8.1x10^{-5}$) in EUR. In AFR, CAD PRS is associated with eye inflammation (OR=0.76, $p=6.6x10^{-5}$).

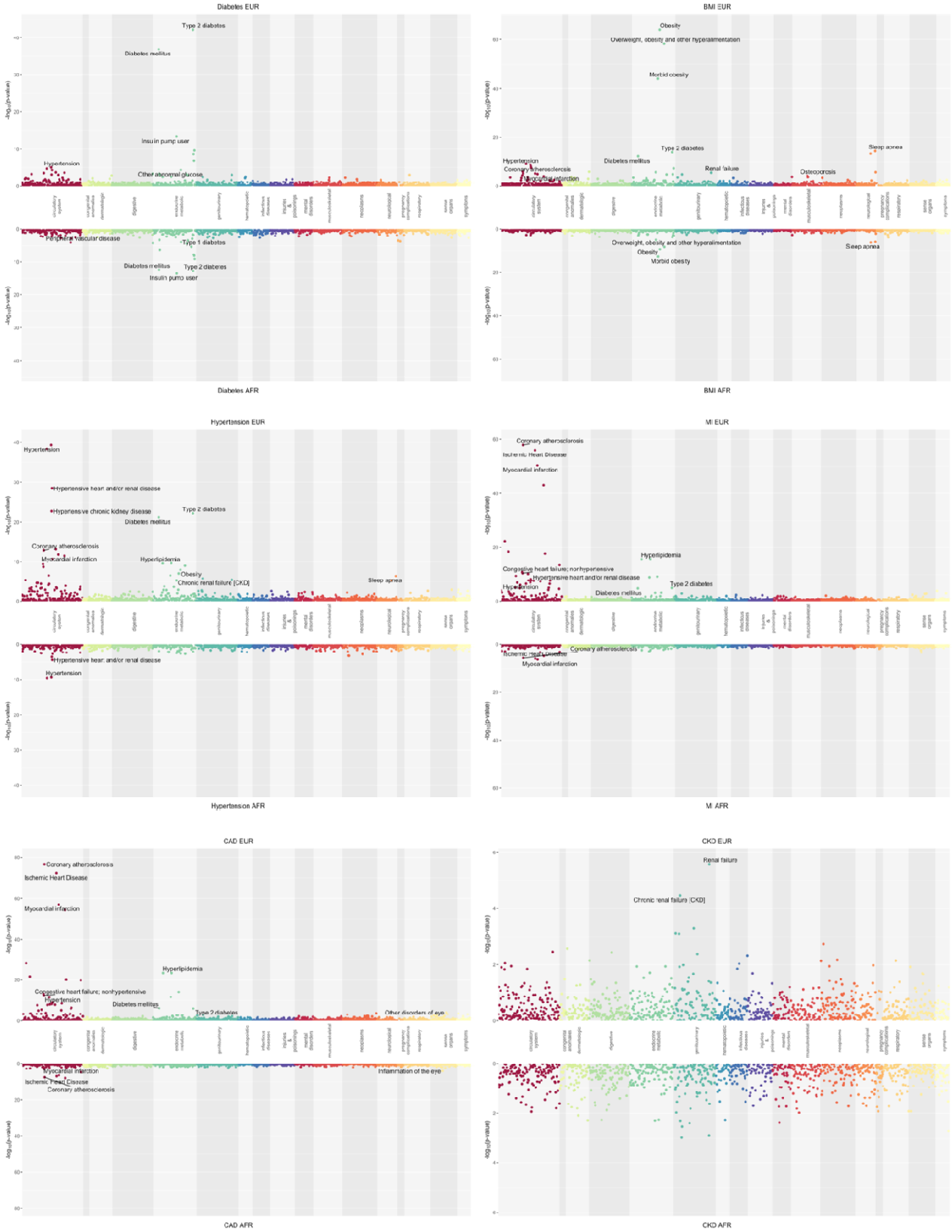

Figure 4. Secondary associations of polygenic risk for T2D, BMI, Hypertension, MI, CAD and CKD in EUR (top panels) and AFR (bottom panels) individuals.

Discussion

We generated six polygenic risk scores representing genetic liability for cardio-renal-metabolic diseases, and investigated their performance in both European and African ancestry individuals in the Penn Medicine BioBank (PMBB), a biobank linked with electronic health records. For all phenotypes tested, we identified a significant association between the PRS and the primary phenotype in both ancestry groups. However, the ability of the PRS to discriminate between cases and controls varied among phenotypes. Furthermore, none of the PRS in AFR were sufficient to stratify individuals according to risk.

In European ancestry individuals, the PRS with the largest effect size was myocardial infarction, followed by coronary artery disease. The two GWAS that were used to generate these PRS also had the largest number of cases, and the PMBB dataset also contained a large number of cases for both of these diseases. However, the CKD PRS was the weakest performer in terms of effect size, despite it being based on the next largest GWAS and the PMBB containing a large number of individuals with renal failure. Therefore, while case number in both the GWAS and the target sample are clearly important, we believe that other factors such as disease heterogeneity, prevalence, penetrance, and non-additive effects among others must also play a role in the ability of PRS to associate with disease.

We conducted a number of analyses to explore secondary phenotypes associated with each PRS. First, we show that the PRS generated for the six diseases are correlated with each other, a finding supported by prior studies showing genetic correlations between CaReMe conditions[4]. We next show that increased PRS is associated with increased burden of disease, suggesting that a higher PRS burden may contribute in a non-discriminating fashion to disease outcome. Finally, we perform PheWAS analysis to identify secondary phenotypes associated with genetic liability for CaReMe diseases. Many of the secondary phenotypes identified could be attributed to the broader effects of disease risk factors and known comorbidities. For instance, risk for Type 2 diabetes was associated with hypertension, a known commonly co-occurring trait[25]. The BMI PRS was associated with sleep apnea, diabetes, hypertension and osteoporosis; all traits known to increase in individuals with higher BMI[26–29]. The extent to which these

secondary phenotypes reflect causal associations between genetic risk and disease is unclear due to the commonality of co-morbidity of these traits.

Our findings highlight a major issue in the future implementation of PRS in clinical care. While GWAS conducted in EUR populations can be used to generate PRS that are associated with phenotype in AFR individuals, the scores generated are not sufficient to differentiate between cases and controls in a predictive model. This was an expected finding, and has been discussed widely in recent years as being a critical source of disparity in genetic research[14,16]. Due to differences in linkage disequilibrium patterns, effect sizes, and causal variants themselves, conducting GWAS in populations that are reflective of the patient population are necessary and will need to be prioritized in the coming years.

Finally, while there is much excitement and enthusiasm about PRS for clinical care, there is still significant research to be conducted to determine its optimal implementation. One of the most essential needs is to investigate how PRS can be incorporated alongside information commonly used to predict patients' risk, such as family history, clinical comorbidities, and environmental/lifestyle factors. Many chronic diseases have published clinical guidelines with risk reduction recommendations (for example CVD[30]). The ultimate clinical utility of PRS will come to fruition when we understand how to integrate PRS with these published guidelines.

Supplemental Information

Supplemental Table 1: Encounter Type (see excel file)

Supplemental Table 2: Association of parameters of each PRS with their primary phenotype in EUR and AFR ancestry.

Disease/PRS	**Parameter**	**EUR**		**AFR**	
		OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**
Type 2 diabetes/ Type 2 diabetes	ρ=0.001	1.09 (1.04-1.15)	$4.73x10^{-4}$	1.13 (1.06-1.21)	$1.57x10^{-4}$
	ρ=0.003	1.1 (1.04-1.15)	$2.33x10^{-4}$	1.19 (1.11-1.27)	$2.85x10^{-7}$
	ρ=0.01	**1.52 (1.43-1.61)**	**$6.62x10^{-43}$**	**1.3 (1.21-1.39)**	**$2.19x10^{-13}$**
	ρ=0.03	1.48 (1.39-1.57)	$2.76x10^{-34}$	1.26 (1.18-1.36)	$1.94x10^{-10}$
	ρ=0.1	1.43 (1.34-1.52)	$5.69x10^{-27}$	1.23 (1.14-1.32)	$6.25x10^{-8}$
	ρ=0.3	1.4 (1.31-1.49)	$1.37x10^{-23}$	1.2 (1.12-1.29)	$1.30x10^{-6}$
	ρ=1	1.38 (1.3-1.48)	$2.02x10^{-22}$	1.19 (1.11-1.28)	$3.83x10^{-6}$
Obesity/ BMI	ρ=0.001	1.18 (1.11-1.24)	$2.81x10^{-9}$	1.07 (1.01-1.13)	0.022
	ρ=0.003	1.17 (1.11-1.24)	$5.22x10^{-9}$	1.04 (0.98-1.1)	0.222
	ρ=0.01	1.55 (1.47-1.64)	$6.85x10^{-55}$	1.06 (1-1.12)	0.059

	ρ=0.03	1.62 (1.53-1.71)	$6.77x10^{-62}$	1.21 (1.14-1.29)	$3.58x10^{-10}$
	ρ=0.1	1.65 (1.56-1.75)	$2.95x10^{-64}$	**1.22 (1.14-1.29)**	**$5.55x10^{-10}$**
	ρ=0.3	**1.67 (1.57-1.77)**	**$8.97x10^{-65}$**	1.21 (1.14-1.29)	$1.53x10^{-9}$
	ρ=1	1.66 (1.57-1.76)	$2.76x10^{-63}$	1.21 (1.13-1.29)	$5.17x10^{-9}$
Hypertension/ Hypertension	ρ=0.001	1.11 (1.06-1.16)	$2.21x10^{-5}$	1.18 (1.1-1.27)	$1.08x10^{-5}$
	ρ=0.003	1.09 (1.04-1.14)	$6.03x10^{-4}$	1.17 (1.09-1.26)	$3.45x10^{-5}$
	ρ=0.01	1.10 (1.05-1.16)	$8.96x10^{-5}$	1.20 (1.11-1.29)	$1.93x10^{-6}$
	ρ=0.03	1.13 (1.07-1.18)	$1.27x10^{-6}$	1.26 (1.17-1.36)	$7.27x10^{-10}$
	ρ=0.1	1.22 (1.16-1.28)	$6.43x10^{-16}$	1.22 (1.13-1.31)	$1.21x10^{-7}$
	ρ=0.3	1.32 (1.26-1.39)	$2.85x10^{-29}$	**1.27 (1.18-1.36)**	**$4.31x10^{-10}$**
	ρ=1	**1.40 (1.33-1.47)**	**$4.50x10^{-40}$**	1.25 (1.16-1.35)	$1.94x10^{-9}$
Myocardial Infarction/ Myocardial Infarction	ρ=0.001	1.64 (1.53-1.77)	$2.22x10^{-41}$	1.23 (1.1-1.37)	$1.39x10^{-4}$
	ρ=0.003	1.74 (1.62-1.88)	$2.02x10^{-49}$	1.26 (1.13-1.4)	$2.13x10^{-5}$
	ρ=0.01	**1.78 (1.65-1.91)**	**$3.74x10^{-51}$**	1.27 (1.14-1.41)	$9.88x10^{-6}$
	ρ=0.03	1.73 (1.6-1.86)	$1.05x10^{-45}$	1.3 (1.17-1.45)	$8.88x10^{-7}$

	ρ=0.1	1.67 (1.55-1.8)	4.29×10^{-40}	**1.32 (1.18-1.47)**	**5.01×10^{-7}**
	ρ=0.3	1.63 (1.51-1.76)	7.15×10^{-37}	1.32 (1.18-1.47)	7.38×10^{-7}
	ρ=1	1.62 (1.5-1.74)	1.77×10^{-35}	1.31 (1.18-1.46)	9.68×10^{-7}
Coronary atherosclerosis/ coronary artery disease	ρ=0.001	1.52 (1.45-1.6)	1.44×10^{-60}	1.17 (1.08-1.27)	2.59×10^{-4}
	ρ=0.003	1.59 (1.51-1.68)	6.12×10^{-71}	1.14 (1.05-1.24)	0.003
	ρ=0.01	**1.66 (1.57-1.75)**	**1.54×10^{-77}**	1.27 (1.17-1.38)	2.87×10^{-8}
	ρ=0.03	1.62 (1.54-1.72)	4.53×10^{-69}	**1.27 (1.17-1.38)**	**2.53×10^{-8}**
	ρ=0.1	1.57 (1.49-1.66)	8.72×10^{-59}	1.25 (1.15-1.36)	1.89×10^{-7}
	ρ=0.3	1.55 (1.46-1.63)	6.01×10^{-54}	1.24 (1.14-1.34)	5.25×10^{-7}
	ρ=1	1.53 (1.45-1.62)	1.04×10^{-51}	1.23 (1.13-1.34)	9.05×10^{-7}
Renal failure/ chronic kidney disease	ρ=0.001	1 (0.95-1.05)	0.935	**1.08 (1.01-1.16)**	**0.024**
	ρ=0.003	1.02 (0.97-1.07)	0.408	1.06 (0.99-1.13)	0.106
	ρ=0.01	**1.16 (1.09-1.24)**	**2.60×10^{-6}**	1 (0.93-1.07)	0.908
	ρ=0.03	1.16 (1.08-1.24)	1.91×10^{-5}	1.05 (0.97-1.14)	0.216
	ρ=0.1	1.14 (1.07-1.23)	1.48×10^{-4}	1.05 (0.96-1.14)	0.285

	ρ=0.3	1.13 (1.06-1.22)	3.99×10^{-4}	1.04 (0.96-1.13)	0.356
	ρ=1	1.13 (1.05-1.21)	5.61×10^{-4}	1.04 (0.95-1.13)	0.379

Supplemental Table 3: AUC of parameters of each PRS with their primary phenotype in EUR and AFR ancestry.

Disease/PRS	**Parameter**	**EUR**			**AFR**		
		AUC (full model)	**AUC (covariates only)**	**AUC difference**	**AUC (full model)**	**AUC (covariates only)**	**AUC difference**
Type 2 diabetes/ Type 2 diabetes	ρ=0.001	0.5764	0.5716	0.0048	0.7046	0.7018	0.0028
	ρ=0.003	0.5769		0.0053	0.7067		0.0049
	ρ=0.01	**0.6218**		0.0502	**0.7116**		0.0098
	ρ=0.03	0.6133		0.0417	0.709		0.0072
	ρ=0.1	0.6063		0.0347	0.7071		0.0053
	ρ=0.3	0.6028		0.0312	0.706		0.0042
	ρ=1	0.6015		0.0299	0.7056		0.0038
Obesity/ BMI	ρ=0.001	0.6006	0.5931	0.0075	0.5976	0.5949	0.0027
	ρ=0.003	0.5991		0.006	0.5959		0.001

	ρ=0.01	0.6453		0.0522	0.5981		0.0032
	ρ=0.03	0.6515		0.0584	0.6119		0.017
	ρ=0.1	0.6539		0.0608	**0.6119**		0.017
	ρ=0.3	**0.6543**		0.0612	0.611		0.0161
	ρ=1	0.6529		0.0598	0.6099		0.015
Hypertension / Hypertension	ρ=0.001	0.6979		0.0017	0.8418		0.0015
	ρ=0.003	0.6978		0.0016	0.8416		0.0013
	ρ=0.01	0.6978		0.0016	0.8421		0.0018
	ρ=0.03	0.698	0.6962	0.0018	0.8434	0.8403	0.0031
	ρ=0.1	0.7031		0.0069	0.8425		0.0022
	ρ=0.3	0.7096		0.0134	**0.8435**		0.0032
	ρ=1	**0.7143**		0.0181	0.8434		0.0031
Myocardial Infarction/ Myocardial Infarction	ρ=0.001	0.7686		0.0279	0.8355		0.0029
	ρ=0.003	0.7738	0.7407	0.0331	0.8359	0.8326	0.0033
	ρ=0.01	**0.7753**		0.0346	0.8361		0.0035

	ρ=0.03	0.7721		0.0314	0.8369		0.0043
	ρ=0.1	0.7684		0.0277	**0.8372**		0.0046
	ρ=0.3	0.7662		0.0255	0.8371		0.0045
	ρ=1	0.7652		0.0245	0.8369		0.0043
Coronary atherosclerosis/ coronary artery disease	ρ=0.001	0.7756		0.018	0.8425		0.0013
	ρ=0.003	0.7793		0.0217	0.8424		0.0012
	ρ=0.01	**0.7818**		0.0242	0.8448		0.0036
	ρ=0.03	0.7794	0.7576	0.0218	**0.8448**	0.8412	0.0036
	ρ=0.1	0.7764		0.0188	0.8443		0.0031
	ρ=0.3	0.7749		0.0173	0.8441		0.0029
	ρ=1	0.7742		0.0166	0.8439		0.0027
Renal failure/ chronic kidney disease	ρ=0.001	0.6271		0	**0.7663**		0.0006
	ρ=0.003	0.6273	0.6271	0.0002	0.766	0.7657	0.0003
	ρ=0.01	**0.6304**		0.0033	0.7657		0
	ρ=0.03	0.6299		0.0028	0.766		0.0003

	ρ=0.1	0.6294		0.0023	0.7659		0.0002
	ρ=0.3	0.6291		0.002	0.7658		0.0001
	ρ=1	0.629		0.0019	0.7658		0.0001

Supplemental Table 4: Step-wise regression for model selection. Using the step-wise regression we added each covariate to the null model and colored cells in the table below show covariates selected in best performing regression model. Best model was selected using akaike information criterion (AIC).

Disease/PRS	Race	PRS	AGE	SEX	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Type 2 diabetes/Type 2 diabetes	EUR													
	AFR													
Obesity/BMI	EUR													
	AFR													
Hypertension/	EUR													

Hypertension	AFR													
Myocardial Infarction/ Myocardial Infarction	EUR													
	AFR													
Coronary atherosclerosis/ coronary artery disease	EUR													
	AFR													
Renal failure/ chronic kidney disease	EUR													
	AFR													

Supplemental Table 5: Correlations between PRS.

Disease 1	Disease 2	EUR		AFR	
		cor	p	cor	p
T2D	BMI	0.216	$<2.2\times10^{-16}$	0.170	$<2.2\times10^{-16}$
T2D	Hypertension	0.049	7.39×10^{-7}	0.057	2.40×10^{-5}
BMI	Hypertension	0.143	$<2.2\times10^{-16}$	0.129	$<2.2\times10^{-16}$
T2D	MI	0.018	0.063	0.095	1.57×10^{-12}
BMI	MI	0.044	9.36×10^{-6}	0.110	4.44×10^{-16}

Hypertension	MI	0.155	$<2.2x10^{-16}$	0.135	$<2.2x10^{-16}$
T2D	CAD	-0.011	0.270	0.033	0.013
BMI	CAD	0.055	$1.85x10^{-8}$	0.055	$4.91x10^{-5}$
Hypertension	CAD	0.179	$<2.2x10^{-16}$	0.136	$<2.2x10^{-16}$
MI	CAD	0.880	$<2.2x10^{-16}$	0.805	$<2.2x10^{-16}$
T2D	CKD	0.315	$<2.2x10^{-16}$	-0.003	0.829
BMI	CKD	0.157	$<2.2x10^{-16}$	0.069	$3.29x10^{-7}$
Hypertension	CKD	-0.022	0.029	-0.006	0.656
MI	CKD	-0.083	$<2.2x10^{-16}$	-0.009	0.489
CAD	CKD	-0.153	$<2.2x10^{-16}$	-0.010	0.451

Supplemental Table 6: PheWAS of T2D PRS (see excel file)

Supplemental Table 7: PheWAS of BMI PRS (see excel file)

Supplemental Table 8: PheWAS of Hypertension PRS (see excel file)

Supplemental Table 9: PheWAS of MI PRS (see excel file)

Supplemental Table 10: PheWAS of CAD PRS (see excel file)

Supplemental Table 11: PheWAS of CKD PRS (see excel file)

References

1. Abraham, G. *et al.* Genomic prediction of coronary heart disease. *Eur. Heart J.* **37**, 3267–3278 (2016).
2. Tada, H. *et al.* Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur. Heart J.* **37**, 561–567 (2016).
3. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

4. Arnold, S. V. *et al.* Burden of cardio-renal-metabolic conditions in adults with type 2 diabetes within the Diabetes Collaborative Registry. *Diabetes Obes. Metab.* **20**, 2000–2003 (2018).
5. Wang, H. *et al.* Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* **388**, 1459–1544 (2016).
6. Ogurtsova, K. *et al.* IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res. Clin. Pract.* **128**, 40–50 (2017).
7. Ma, R. C. Genetics of cardiovascular and renal complications in diabetes. *J. Diabetes Investig.* **7**, 139–154 (2016).
8. Regele, F. *et al.* Genome-wide studies to identify risk factors for kidney disease with a focus on patients with diabetes. *Nephrol. Dial. Transplant.* **30**, iv26–iv34 (2015).
9. Rangaswami, J. *et al.* Cardiorenal Syndrome: Classification, Pathophysiology, Diagnosis, and Treatment Strategies: A Scientific Statement From the American Heart Association. *Circulation* **139**, (2019).
10. Denny, J. C. *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
11. Verma, A. & Ritchie, M. D. Current Scope and Challenges in Phenome-Wide Association Studies. *Current Epidemiology Reports* (2017). doi:10.1007/s40471-017-0127-7
12. Bush, W. S., Oetjens, M. T. & Crawford, D. C. Unravelling the human genome–phenome relationship using phenome-wide association studies. *Nat. Rev. Genet.* **17**, 129–145 (2016).
13. Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J. & Lachance, J. Genetic disease risks can be misestimated across global populations. *Genome Biol.* **19**, (2018).
14. De La Vega, F. M. & Bustamante, C. D. Polygenic risk scores: a biased prediction? *Genome Med.* **10**, (2018).
15. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
16. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
17. Gurdasani, D., Barroso, I., Zeggini, E. & Sandhu, M. S. Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* (2019). doi:10.1038/s41576-019-0144-0
18. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
19. Verma, S. S. *et al. Imputation and quality control steps for combining multiple genome-wide datasets*.

20. Galinsky, K. J. *et al.* Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016).
21. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
22. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, (2015).
23. Denny, J. C. *et al.* Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1110 (2013).
24. Wu, P. *et al.* Developing and Evaluating Mappings of ICD-10 and ICD-10-CM Codes to PheCodes. *bioRxiv* (2019). doi:10.1101/462077
25. Sun, D. *et al.* Type 2 Diabetes and Hypertension: A Study on Bidirectional Causality. *Circ. Res.* **124**, 930–937 (2019).
26. Romero-Corral, A., Caples, S. M., Lopez-Jimenez, F. & Somers, V. K. Interactions Between Obesity and Obstructive Sleep Apnea. *Chest* **137**, 711–719 (2010).
27. Dua, S., Bhuker, M., Sharma, P., Dhall, M. & Kapoor, S. Body mass index relates to blood pressure among adults. *North Am. J. Med. Sci.* **6**, 89 (2014).
28. Xiang, B.-Y. *et al.* Body mass index and the risk of low bone mass–related fractures in women compared with men: A PRISMA-compliant meta-analysis of prospective cohort studies. *Medicine (Baltimore)* **96**, e5290 (2017).
29. Gray, N., Picone, G., Sloan, F. & Yashkin, A. Relation between BMI and Diabetes Mellitus and Its Complications among US Older Adults. *South. Med. J.* **108**, 29–36 (2015).
30. Arnett, D. K. *et al.* 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* (2019). doi:10.1161/CIR.0000000000000678